

## Article

# Genome-Wide Identification and Characterization of the *PPPDE* Gene Family in Rice

Wangmin Lian <sup>1,†</sup>, Xiaodeng Zhan <sup>1,†</sup>, Daibo Chen <sup>1</sup>, Weixun Wu <sup>1</sup>, Qunen Liu <sup>1</sup>, Yinxing Zhang <sup>1</sup>, Shihua Cheng <sup>1</sup>, Xiangyang Lou <sup>1</sup>, Liyong Cao <sup>1,2</sup> and Yongbo Hong <sup>1,3,\*</sup>

- <sup>1</sup> China National Center for Rice Improvement, State Key Laboratory of Rice Biology and Breeding, Zhejiang Key Laboratory of Super-Rice, China National Rice Research Institute, Hangzhou 311401, China; lianwangm@gmail.com (W.L.); zhanxiaodeng@caas.cn (X.Z.); chendaibo@caas.cn (D.C.); wuweixun@caas.cn (W.W.); liuqunen@caas.cn (Q.L.); zhangyingxin@caas.cn (Y.Z.); chengshihua@caas.cn (S.C.); louxiangyang01@caas.cn (X.L.); caoliyong@caas.cn (L.C.)
- <sup>2</sup> Northern Rice Center, China National Rice Research Institute, Shuangyashan 155600, China
- <sup>3</sup> National Nanfan Research Institute, Chinese Academy of Agricultural Sciences, Sanya 572025, China
- \* Correspondence: hongyongbo@caas.cn; Tel.: +86-0571-63370338
- † These authors contributed equally to this work.

**Abstract:** Protein ubiquitination is common and crucial in cellular functions, however, little is known about how deubiquitinating enzymes (DUBs) reverse regulate the ubiquitination signaling process. *PPPDE* family proteins are a novel class of deubiquitinating peptidases with demonstrated deubiquitination/deSUMOylating activities. In this study, we identified 10 *PPPDE* genes from the rice (*Oryza sativa* L.) genome unevenly distributed on five chromosomes, where most of these members have not been reported to date. Based on the gene structure, the *OsPPPDE* family consists of three distinct subgroups within the phylogenetic tree. *Cis*-element analysis identified light/phytohormone-responsive, development, and abiotic stress-related elements in the promoters of *OsPPPDE*. Furthermore, we conducted and analyzed the transcript abundance of *OsPPPDE* under various tissues and stresses using the transcriptome data of 352 samples from the Rice Expression Database and GEO datasets. Moreover, *OsPPPDE5* showed differential regulation of its transcript abundance during Cd and drought stress. Collinearity and syntenic analysis of 101 *PPPDEs* and *PPPDE*-like proteins in 10 plant genomes indicated that this family is evolutionarily conserved. Domestication analysis suggests that *OsPPPDEs* may contribute to *indica-japonica* divergence using the data from the 3K Rice Genome Project. Our study provides a foundation for further study on the function and molecular mechanism of the *OsPPPDE* gene family.

**Keywords:** *PPPDE* gene family; rice; phylogenetic analysis; domestication; stress response



**Citation:** Lian, W.; Zhan, X.; Chen, D.; Wu, W.; Liu, Q.; Zhang, Y.; Cheng, S.; Lou, X.; Cao, L.; Hong, Y.

Genome-Wide Identification and Characterization of the *PPPDE* Gene Family in Rice. *Agronomy* **2024**, *14*, 1035. <https://doi.org/10.3390/agronomy14051035>

Received: 13 March 2024

Revised: 26 April 2024

Accepted: 2 May 2024

Published: 13 May 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The ubiquitin signaling network plays major roles in the regulation of cell cycle progression including RNA metabolism, pathogen response, and programmed cell death (apoptosis) in eukaryotic cells [1,2]. The ubiquitin signaling network is assembled by a triad of ubiquitin (Ub), ubiquitin-activating (E1), ubiquitin-conjugating (E2), ubiquitin ligase (E3), 26S proteasome (26S), and deubiquitinating enzymes (DUBs) [3,4]. Ubiquitination regulates subcellular localization, enzyme activity, and protein interactions, but target protein multiubiquitination is usually the signal for the ubiquitin hydrolysis of proteins [5]. DUBs that can remove ubiquitin tags from substrate proteins are an essential regulatory factor in the Ub-dependent pathway [6]. DUBs have four molecular roles: (1) Processing Ub precursors; (2) generating free Ub monomers from Ub precursors and Ub adducts as well as from the poly-Ub chains produced [7]; (3) recycling of Ub or Ub-oligomers from Ub-protein conjugates targeted for degradation [8]; and (4) preventing the degradation of pre-targeted proteins by releasing intact ubiquitin and targets to conjugates [9]. DUBs are carried out

by a classical evolutionary conserved deubiquitinase superfamily known for countering the action of E3 ligases. In plants, all seven known DUB subfamilies have been identified [10], namely Ub-binding protease/Ub-specific protease (UBP/USP), Ub C-terminal hydrolase (UCH), Machado–Joseph domain-containing protease (MJD), ovarian-tumor domain-containing protease (OTU), zinc finger with UFM1-specific peptidase domain protease (ZUFSP), motif interacting with Ub-containing novel DUB family (MINDY), and the JAB1/MPN/MOV34 protease (JAMM) [11–13]. The regulation of ubiquitylation is the key to plant growth and development, in which the activities of ubiquitylating enzymes as well as DUBs determine the stability or function of the modified proteins [14]. UBP24 may have a crucial role in responding abscisic acid (ABA) signaling and the drought and salt of *Arabidopsis thaliana* as *ubp24* mutants have been shown to be hypersensitive to ABA and salt stress. In addition, UBP24 can form homodimers and cleave UBQ1 into Ub-monomers [15]. UBP8, in *Magnaporthe oryzae*, is involved in carbon catabolite repression, appressoria development, melanin formation, and the pathogenicity of *M. oryzae* [6]. OTLD1, which belongs to the OTU (ovarian tumor protease) deubiquitinase families, increases the cell size and stimulates plant growth due to the transcriptional repression of multiple major genes of *Arabidopsis* growth and development [16].

The PPPDE superfamily (after Permuted Papain fold Peptidases of DsRNA viruses and Eukaryotes) has been identified as putative deubiquitinating isopeptidases and is involved in de-Ub and/or deSUMOylation in most eukaryotic lineages [17]. The PPPDE domain consists of thiol peptidases with a circularly permuted papain-like fold. The secondary structure of PPPDE, which has an  $\alpha + \beta$  fold with five conserved  $\alpha$ -helices and six  $\beta$ -strands with the conserved histidine positioned at the N-terminus of strand 2, is similar to the classic papain-like proteases. PPPDE superfamily proteins have a strictly conserved polar residue at the C-terminus to orient the catalytic histidine [17,18]. The PPPDE superfamily is not only found in eukaryotes, but also in dsRNA viruses and a single-stranded DNA virus. Furthermore, phylogenetic tree analysis showed that PPPDE proteins can be divided into three families, in particular, the family 1 PPPDEs, which can be traced back to the last common ancestor of eukaryotes. The presence of PPPDE protein in the early stage of eukaryotic evolution suggests that it may target evolutionarily conserved proteins, and join in the regulation of some crucial cellular processes like other DUBs [17].

Although some researchers think that the function of the PPPDE domain might be partially redundant with that of other DUBs due to repeated gene loss events in evolution, recent studies have indicated that PPPDE is involved in cell cycle regulation, embryonic development, and apoptosis. DeSUMOylating isopeptidase 1 (DeSI1; previously described as PPPDE2), a member of the human and mouse PPPDE peptidase domain superfamily, catalyzed the deSUMOylation of BZEL [19]. In addition, PPPDE1 also has deubiquitinating and deSUMOylating activity [20]. The substrate proteins of PPPDE1 are ribosomal proteins S7 and RPS7, and PPPDE1 could mediate the ubiquitin chain editing of RPS7, stabilizing the RPS7 protein [21,22]. The catalytic domains of animal and plant DUB families show high homology, whereas the regions outside of the catalytic site can vary significantly [14]. In *Arabidopsis*, *Desi1* and *AtC3H59* interact in the nucleus to jointly regulate seed germination, seedling development, and seed development [23]. Although a recent study has reported that PIC11 stabilizes OsMETS to promote Met–ethylene biosynthesis, which contributes to broad-spectrum blast resistance [24], the function of OsPPPDE superfamily has still not been reported yet, except for PIC11. The objective of our research was to identify PPPDE genes and investigate the evolutionary pattern and functions of the PPPDE superfamily in rice through a whole genome data and annotation analysis.

## 2. Materials and Methods

### 2.1. Identification of OsPPPDE Amino Acid Sequences

Whole genome data and annotation of *Oryza sativa* Nipponbare were downloaded from the Rice Genome Annotation Project [25]. Hidden Markov model (HMM) analysis results of the PPPDE putative peptidase domain (NO. PF00145) were downloaded from the

Pfam database [26]. HMMER 3.0 software [27] was employed to search for members of the OsPPPDE proteins with a cut-off E-value of  $1 \times 10^{-5}$ . The obtained proteins were examined for the existence of the PPPDE domain using the SMART online software tool [28]. The molecular weights (MWs) and isoelectric points (pIs) of the members of OsPPPDE proteins were calculated with ExPASy online software [29]. The predicted subcellular localization of proteins were estimated using WolfPsort II [30].

## 2.2. Gene Duplication, Phylogenetic Relationship, Intron–Exon, and Conserved Motifs Analysis

The reference genome data of nine plants were acquired from Ensembl Plants (<https://plants.ensembl.org/index.html>) including *Arabidopsis thaliana*, *Brachypodium distachyon*, *Glycine max*, *Hordeum vulgare*, *Oryza rufipogon*, *Setaria italica*, *Sorghum bicolor*, *Triticum aestivum*, and *Zea mays*. To reveal the collinearity of the OsPPPDE gene family, we used MCScanX to identify homologous gene pairs and syntenic relationships [31]. The location information of all OsPPPDE genes including the exons, introns, and UTR location information on the chromosome was available in the gff3 annotation file of the genome from the Rice Genome Annotation Project [25]. The results of syntenic analysis, chromosomal localization (Figure S1), and gene structure were diagrammed using Tbtools [32]. The phylogenetic tree was built with PPPDE amino acid sequences by MEGA-X software (bootstrap for 1000 replicates) [33]. The result of the phylogenetic tree was pictured using ggtree [34]. The full-length protein sequences of the OsPPPDE were used to search for potential conserved motifs using the MEME [35] online webserver with the default parameters (Figure S2).

## 2.3. Cis-Acting Element Analysis of OsPPPDE Genes Promoter

To search the functional promoter regions, 2 kb sequences located upstream of all OsPPPDE genes were extracted to analyze the *cis*-element of genes. The PlantCARE database [36] was used to predict the *cis*-acting elements, which were visualized using gggenes and ggplot2 [37].

## 2.4. Prediction of Secondary and Tertiary Structures of Proteins

The SOPMA database [38] was used to predict the secondary structures, and the AlphaFold2 database [39] was used to predict the tertiary structures. The PPPDE domain structure model (No. 2WP7) was obtained from the protein structure models of the Protein Data Bank (PDB). The 3D structure analysis of OsPPPDE proteins was conducted using PyMOL [40] (Figure S3).

## 2.5. Expression Analysis of PPPDE in Rice

Gene expression analysis of the rice PPPDE genes was conducted by utilizing the Rice Expression Database (RED) [41] and GEO datasets (GSE118355; GSE126961; GSE81906; GSE81108; GSE84800; GSE131641) and fragments per kilobase millions (FPKM) were downloaded. The rice morphology diagrams and heat map were visualized using gganatogram [42] and ggplot2 [37].

We constructed a co-expression network for the normal type and blast inoculation leaf by calculating the correlation among all rice expressed genes using Spearman's correlation ( $R_s$ ). We adjusted the  $p$  value with the FDR method for multiplicity and considered  $R_s > 0.7$  and  $FDR < 0.01$  as the significance cut-off point. The PCA results (Figure S4) of all 352 samples from the RED and GEO datasets showed that there was not a significant batch effect among the 30 RNA-Seq projects. The detailed experimental information is summarized in Table S1, and the expression abundance data without the batch effect is recorded in Table S2.

## 2.6. Domestication Analysis of OsPPPDE Genes

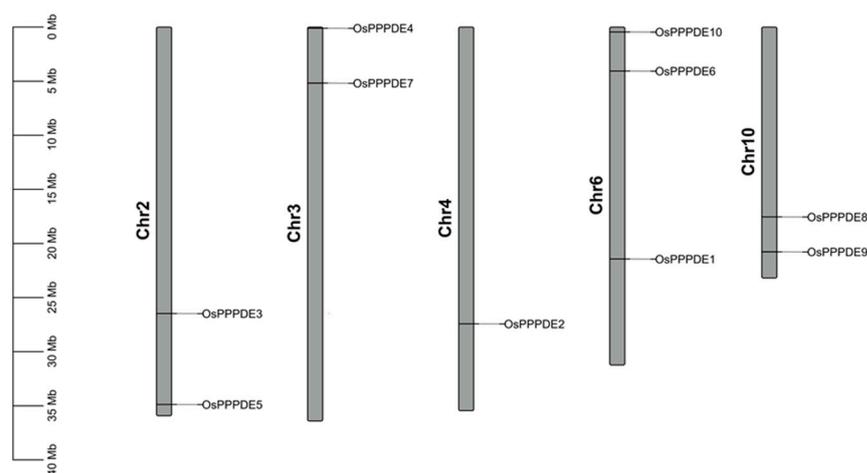
Single-nucleotide polymorphisms (SNPs) that are located at the flanking regions were identified in 3000 Rice Genome Project [43] (Table S3). Fixation index ( $F_{ST}$ ), nucleotide diver-

sity ( $\pi$ ), and Tajima's  $D$  were calculated using a sliding window approach by vcfTools [44]. The distribution of  $F_{ST}$  and  $\pi$  was plotted in sliding windows of 2 kb with a 200-bp step size and the distribution of Tajima's  $D$  was drawn in sliding windows of 2 kb. The linkage disequilibrium (LD) heat map (Figure S5) for *OsPPPDE* gene and flanking regions in five rice groups were visualized LDheatmap [45].

### 3. Results

#### 3.1. Identification of *OsPPPDE* Genes from Rice Genome

To identify the *OsPPPDE* genes, amino acid sequences of the conserved PPPDE domain (PF05903) were examined through the whole rice genome using HMM. As a result, a total of 10 *PPPDE* genes were found; three of them had two transcripts. The *OsPPPDE*s were named as *OsPPPDE1* to *OsPPPDE10* according to their similarities with the PPPDE domain. Next, we collected more information on the *OsPPPDE*s including the chromosomal locations, the length of coding sequence (CDS) and amino acid (aa) sequence, the number of exons, the molecular weight (MW), and the isoelectric point (pI) (Table 1). The encoding protein length of these genes varied from 176 to 315 aa, and there was a big difference in their pIs, which ranged between 4.69 and 10.4. Noticeably, the majority (8/10) of them were annotated as ethylene-responsive element-binding protein, whereas the others were not. The *OsPPPDE*s were located on Chr2, Chr3, Chr4, Chr6, and Chr10. The syntenic relationship analysis showed that *OsPPPDE2* and *OsPPPDE3* were a collinear pair, which were localized on Chr4 and Chr6, respectively (Figure 1).

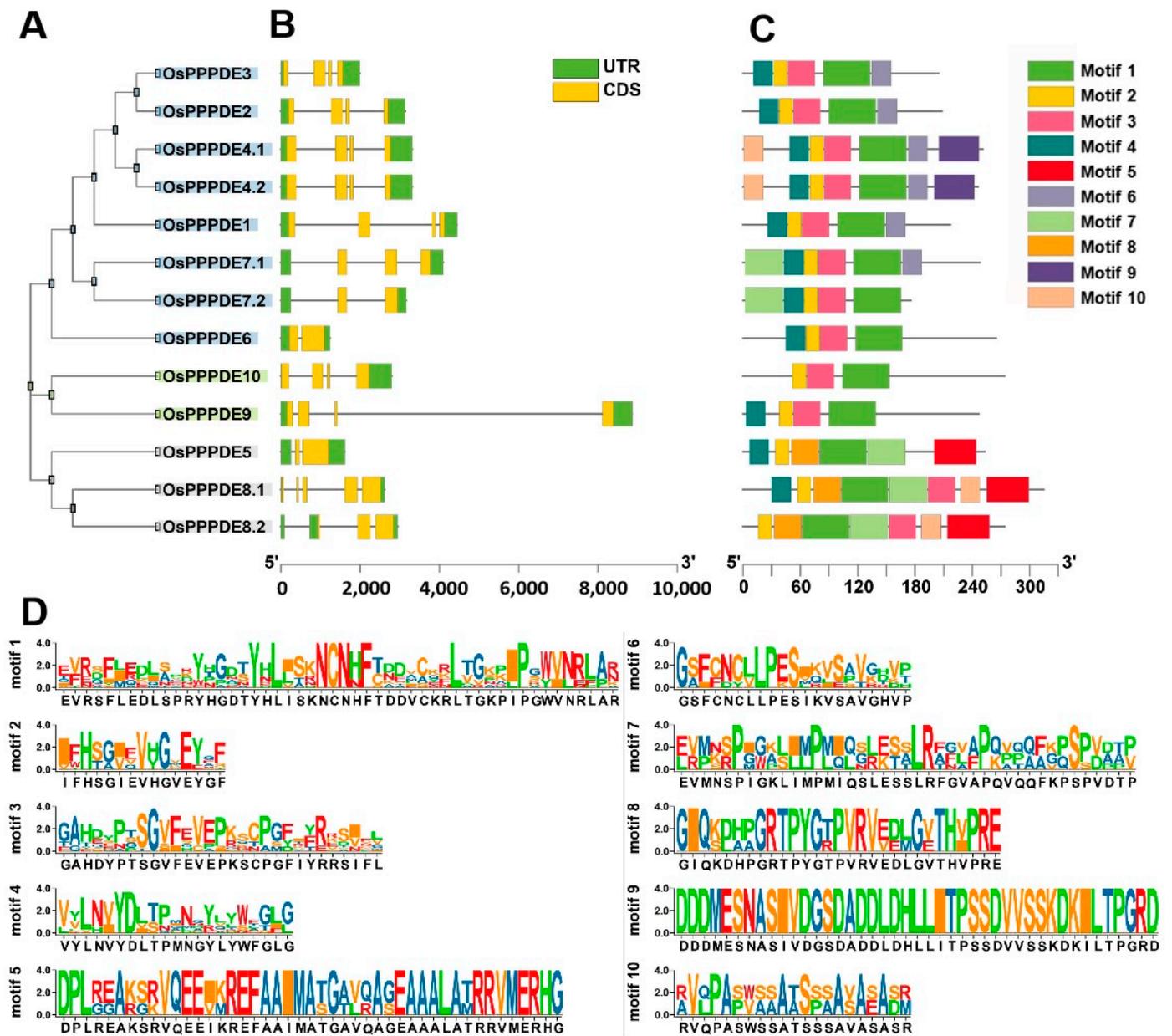


**Figure 1.** Chromosomal localization of the *OsPPPDE* genes in the rice genome. The left is the chromosome number. The right of the chromosomes marked the *OsPPPDE* genes. The scale bar shows the length of the chromosomes.

To deeply understand the evolutionary relationship of *OsPPPDE*s, we built the phylogenetic tree with the maximum likelihood method (ML) using amino acid sequences (Figure 2A). The members of *OsPPPDE*s were grouped into three groups; group I consists of *OsPPPDE3*, *OsPPPDE3*, *OsPPPDE2*, *OsPPPDE4.1*, *OsPPPDE4.2*, *OsPPPDE1*, *OsPPPDE7.1*, *OsPPPDE7.2*, *OsPPPDE6*; group II consists of *OsPPPDE10* and *OsPPPDE9*; and group III consists of *OsPPPDE5*, *OsPPPDE8.1* and *OsPPPDE8.2*.

**Table 1.** Information of the *OsPPPDE* family genes identified in rice genome.

Common Name	RGAP-ID	RAP-ID	Chr	CDS Length (bp)	Exon Number	Protein Length (aa)	Protein MW (Da)	pI	Predicted Subcellular Localization	Annotation	Ref.
OsPPPDE1	LOC_Os06g36490.1	Os06t0560400-01	Chr.6	651	4	217	23,557.16	5.34	cyto, nucl	ethylene-responsive element-binding protein	
OsPPPDE2	LOC_Os04g46290.1	Os04t0548000-01	Chr.4	624	4	208	22,712.3	5.89	cyto, nucl	ethylene-responsive element-binding protein	
OsPPPDE3	LOC_Os02g43840.1	Os02t0655500-01	Chr.2	615	4	205	22,535.32	5.36	cyto, chlo, nucl	ethylene-responsive element-binding protein	
OsPPPDE4.1	LOC_Os03g01130.1	Os03t0100900-01	Chr.3	753	4	251	26,965.1	4.89	mito, chlo, nucl	ethylene-responsive element-binding protein	
OsPPPDE4.2	LOC_Os03g01130.2	Os03t0100900-02	Chr.3	738	4	246	26,491.62	4.97	mito, chlo, vacu	ethylene-responsive element-binding protein	
OsPPPDE5	LOC_Os02g56900.1	Os02t0814000-01	Chr.2	759	3	253	27,286.97	6.12	cyto, chlo, nucl_plas, nucl, plas, mito	thioredoxin family protein	[24]
OsPPPDE6	LOC_Os06g08360.1	Os06t0182100-00	Chr.6	795	2	265	27,149.57	10.4	chlo	ethylene-responsive element-binding protein	
OsPPPDE7.1	LOC_Os03g10200.1	Os03t0198500-01	Chr.3	744	4	248	27,976.24	9.57	nucl, mito, chlo, plas	ethylene-responsive element-binding protein	
OsPPPDE7.2	LOC_Os03g10200.2	Os03t0198500-02	Chr.3	528	3	176	20,110.32	9.44	mito, cyto_mito, chlo, nucl	ethylene-responsive element-binding protein	
OsPPPDE8.1	LOC_Os10g33350.1		Chr.10	945	5	315	33,625.28	8.83	mito, chlo_mito, chlo	endo-beta-N-acetylglucosaminidase	
OsPPPDE8.2	LOC_Os10g33350.3	Os10t0472400-01	Chr.10	822	4	274	28,803.75	6.23	cyto, nucl, cysk_nucl	endo-beta-N-acetylglucosaminidase	
OsPPPDE9	LOC_Os10g38970.1	Os10t0533900-01	Chr.10	741	4	247	27,221.35	6.53	nucl, chlo, cyto, protolasts	ethylene-responsive element-binding protein	[24]
OsPPPDE10	LOC_Os06g01780.1	Os06t0107000-01	Chr.6	822	4	274	29,485.42	4.69	nucl, chlo	ethylene-responsive element-binding protein	

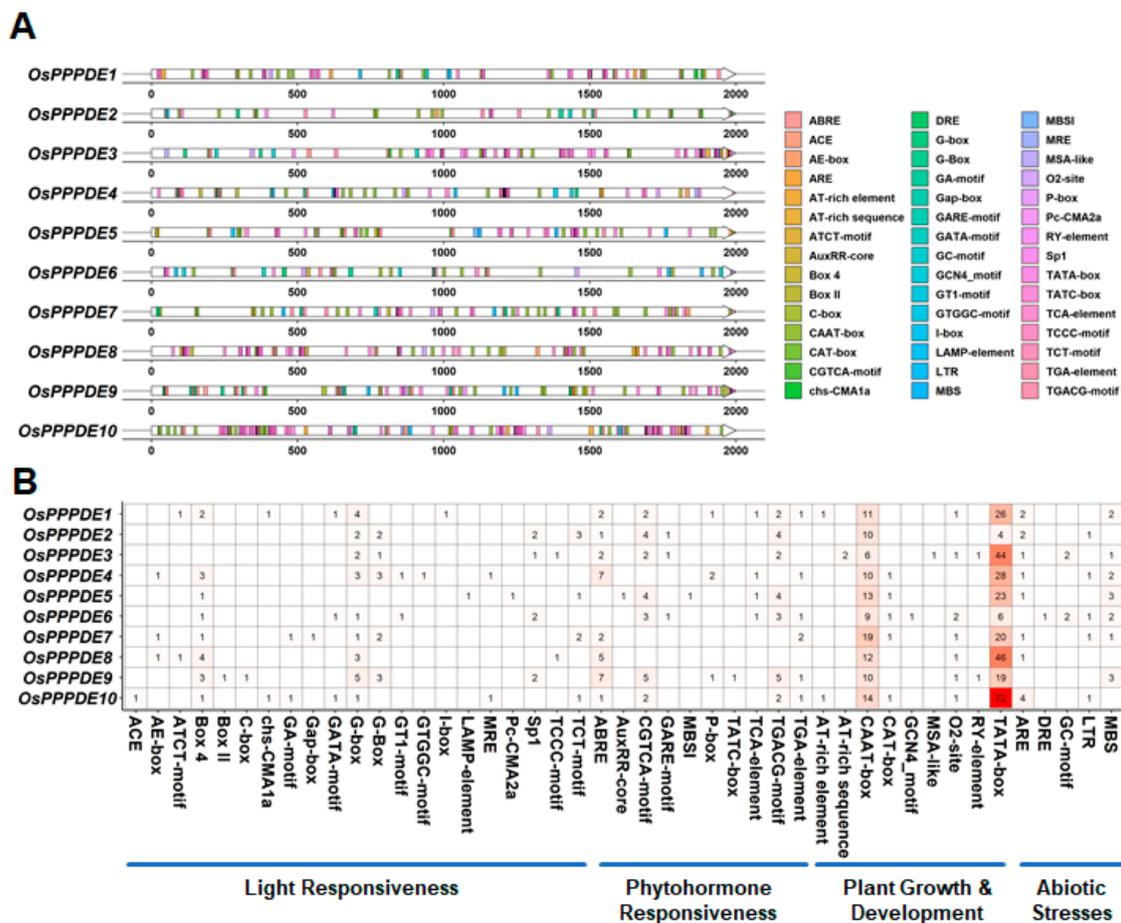


**Figure 2.** Phylogenetic and structural analysis of the *OsPPPDE* genes and encoded proteins in rice. (A) Unrooted phylogenetic tree of *OsPPPDE*s. Group 1, group 2, and group 3 are displayed in blue, green, and grey colors, respectively. (B) Gene structure of the *OsPPPDE*s. Green shows untranslated regions (UTRs); yellow shows coding sequence (CDS); lines show introns. The scale bar estimates the length of the genomic DNA sequence of *OsPPPDE*s at the bottom. (C) Motif distribution of *OsPPPDE*s. Each motif is represented by a specific color. The scale bar at the bottom indicates the length of the motif protein sequence. (D) Motif sequences identified in the *OsPPPDE* proteins. The height of different amino acids represents the repeatability. The scale bar at the bottom represents the amino acid with a large proportion of each motif site.

We compared the DNA sequence of *OsPPPDE*s to analyze the composition of introns and exons; most of them contained two to five exons (Figure 2B), and *OsPPPDE9* had the longest intron length. We also identified 10 conserved motifs (Figure 2C,D). Motif 1, motif 3, and motif 5 were related to the PPPDE domain, but others had no functional annotation, and motif 1 presented in all *OsPPPDE* genes.

### 3.2. Cis-Acting Element Analysis of OsPPPDE Genes Promoter

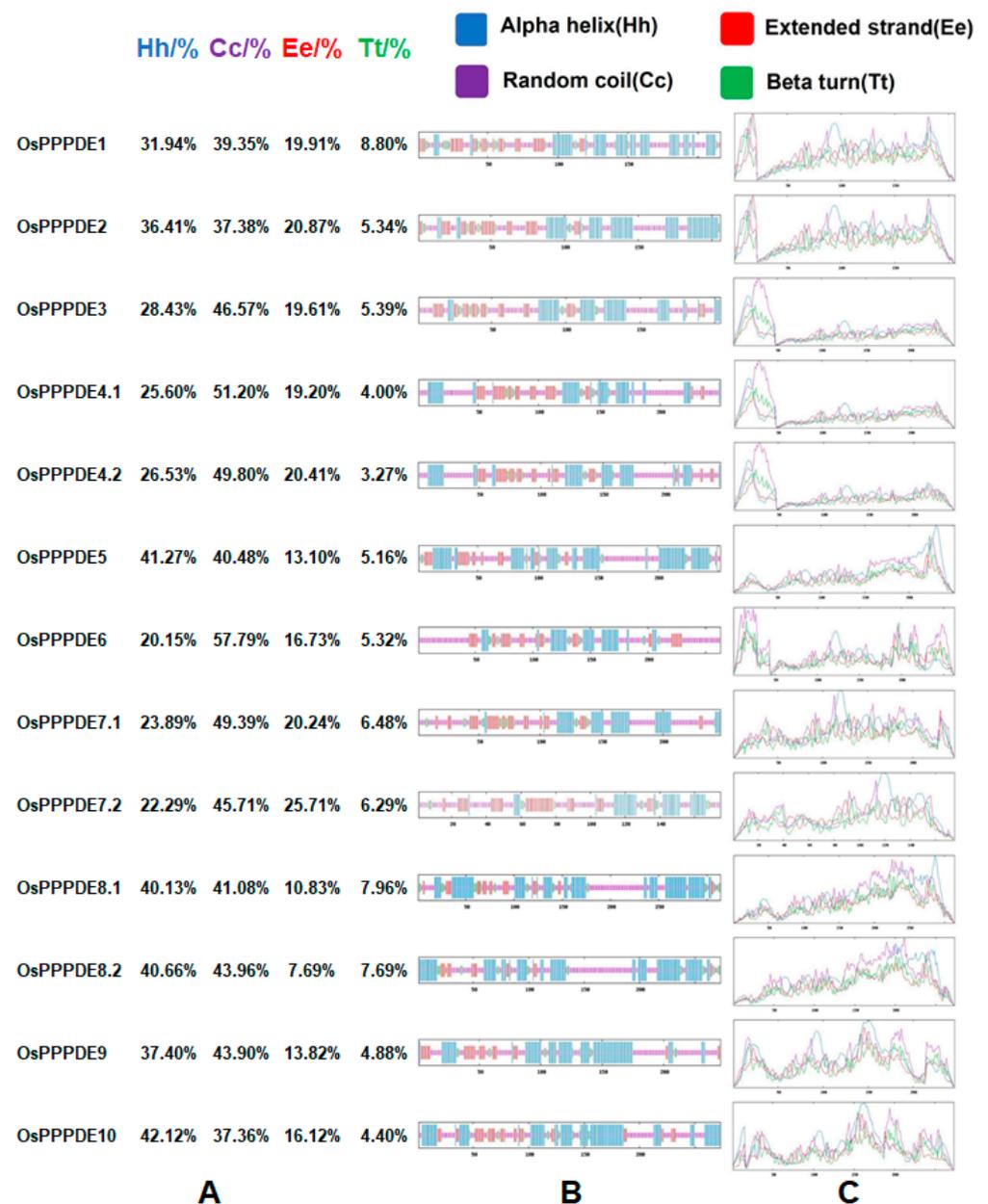
Since *Cis*-acting regulatory elements play a key role in the gene regulation of gene expression patterns in response to hormones and environmental stresses [46], we scanned potential *cis*-acting elements using the PlantCARE database in the upstream regions (2 kb) of *OsPPPDEs* (Figure 3). In total, forty-five *cis*-acting elements were predicated in all *OsPPPDEs*, which could be classified into four groups in light of their putative functions. Light responsive elements included 21 members, which was the largest subdivision including G-box (involved in light responsiveness) and Box 4 (part of a conserved DNA module involved in light responsiveness) as representatives. Regulatory elements related to phytohormone-response comprised 10 components, which involved methyl-jasmonate (MeJA), abscisic acid (ABA), gibberellic acid (GA), salicylic acid (SA), auxin, and flavonoid responsive elements, moreover, abscisic acid responsiveness element (ABRE) was found in eight *OsPPPDEs*. Core promoter elements such as TATA-box and CAAT-box were identified in all *OsPPPDEs*. The number of elements on the promoter of *OsPPPDE10* was the largest, most of which (72/108) were TATA-box; the number of elements on the promoter of *OsPPPDE2* was the least, and the greatest type (10/36) was CAAT-box. Additionally, the stress-related elements were also found in *OsPPPDEs*, these included drought, salinity, low-temp, and anoxic responsive elements. These results suggest that *OsPPPDE* genes are extensively involved in regulating different growth and developmental processes and phytohormone responses in rice.



**Figure 3.** *Cis*-acting element distribution and its function in all *OsPPPDE* genes. (A) Distribution of *cis*-acting elements in upstream (2 kb) of the *OsPPPDE* gene family. The specific colored boxes show different promoter components. (B) Heat diagram of the *cis*-acting element. The numbers and colors in the grid indicate the number of components. The line at the bottom of figure divides elements into light responsiveness, phytohormone responsiveness, plant growth and development, and abiotic stresses.

### 3.3. Prediction of Secondary and Tertiary Structures of OsPPPDEs

Protein structures can provide valuable information that can be used to infer biological processes [47], so we utilized SOPMA and AlphaFold2 to predict secondary structure distributions and the 3D structures of OsPPPDEs. All four secondary structure patterns ( $\alpha$ -helix, random coil, extend strand, and  $\beta$ -turn) were found in the OsPPPDE proteins (Figure 4). The results revealed that all OsPPPDE proteins were mainly comprised of  $\alpha$ -helix and random coil, and the sum of these two patterns accounted for between 68% and 85% of the full length of the protein. Only OsPPPDE5 and OsPPPDE10 had more  $\alpha$ -helix than random coil, the rest had more random coil than  $\alpha$ -helix. Most OsPPPDE proteins consisted of around 20% extended strands and 6%  $\beta$ -turns.



**Figure 4.** Secondary structure of OsPPPDE proteins. (A) Percentage of each secondary structure relative to the full length of the protein. (B) Different rendering types of the secondary structure of OsPPPDE. The specific colors show four types: blue is  $\alpha$ -helix, purple is random coil, red is extended strand, and green is  $\beta$ -turn. (C) Scores of four secondary structures on protein sequences. Color annotations are the same as before.

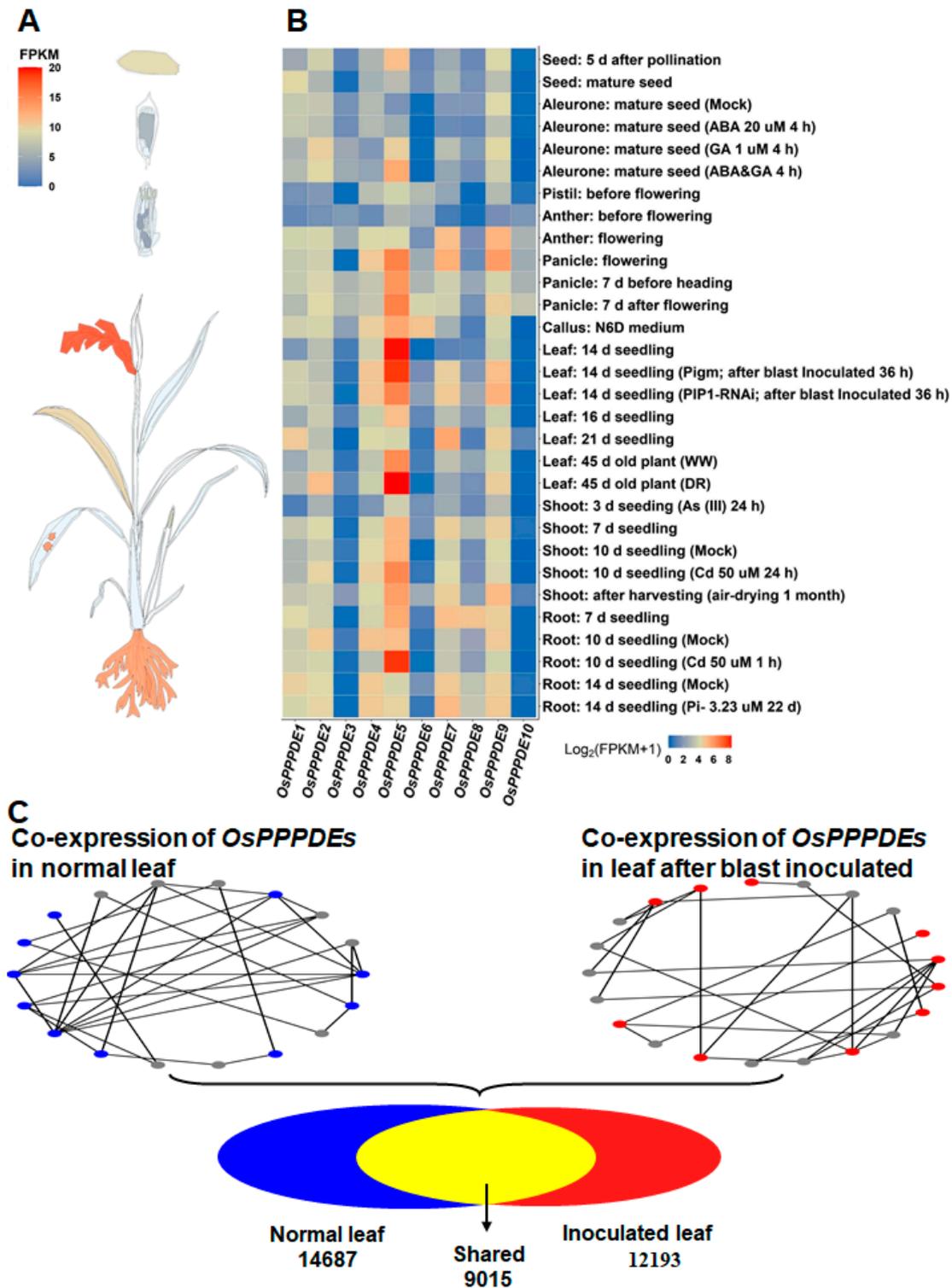
We further predicted the 3D structures of these OsPPPDE proteins using Alphafold2 (neural network based) modeling (Figure S3). The per-residue confidence score (pLDDT) of the PPPDE domain in the OsPPPDE proteins was high, above 90. The PPPDE domain was usually located at the start of the N-terminus of the OsPPPDEs protein, but it was located in the middle (after residue 40th) of OsPPPDE4, OsPPPDE6, and OsPPPDE7. Although structures other than the PPPDE domain could also be predicted such as  $\alpha$ -helix on OsPPPDE5, OsPPPDE8, OsPPPDE9, and OsPPPDE10, most of the remaining pLDDTs were less than 50.

### 3.4. Expression Pattern and Co-Expression Network of OsPPPDEs

To comprehensively investigate the expression at the mRNA level of OsPPPDEs, we collected the RNA-Seq data of 142 normal treatment samples from RED and GEO datasets (see the “Section 2”). The expression levels of the rice PPPDE genes varied significantly in different tissues (Figures 5A and S4B). Among nine tissues, the average OsPPPDE gene expression was the highest in the panicle, followed by the root, and the lowest was in the pistil (Figure 5A). Moreover, the variation in the expression levels was the largest in the leaves for all *OsPPPDE* genes compared to the other tissues (Figure S4B). *OsPPPDE3*, *OsPPPDE6*, and *OsPPPDE10* exhibited a moderate expression level in the anther and panicle, whereas they were negligibly expressed in the leaf, shoot, and root. Additionally, *OsPPPDE5* was expressed highly in most tissues, except aleurone. These tissue expression patterns suggest that *OsPPPDEs* played an influential role in rice development.

Furthermore, the expression patterns of rice PPPDE genes under 30 treatment conditions were also investigated from the RED and GEO datasets (Figure 5B). Results showed that the expressions of OsPPPDEs showed a difference under biotic/abiotic stresses, except for *OsPPPDE3* and *OsPPPDE10*. For instance, *OsPPPDE2*, *OsPPPDE5*, and *OsPPPDE9* had upregulated expression under drought stress. *OsPPPDE5* was highly induced to express in cadmium (Cd)-treated shoots and roots. Both *OsPPPDE7* and *OsPPPDE9* significantly changed after blast inoculation. Furthermore, *OsPPPDE5* was also strongly upregulated by phytohormones like ABA and GA. These results illustrate that the rice PPPDE genes play an important role in response to diverse abiotic stresses and phytohormone responses.

As *OsPPPDE9* (PIC1) contributes to broad-spectrum blast resistance [24], we constructed the co-expression network of PPPDEs in two types of rice leaf, of which one was for normal leaves, and the other was after blast inoculation (Figure 5C). We built the co-expression network by calculating *R*s and FDR among all of the rice expressed genes. There were 14,687 and 12,193 high-confident co-expressed gene pairs identified in the normal treatment leaf and inoculated leaf, respectively. Among them, a total of 9015 gene pairs were shared in two treatments. Taken together, our results suggest functional renewing of the PPPDE network in biotic stress.



**Figure 5.** Global disruption of the *OsPPPDE* co-expression network in rice. (A) The rice morphology diagrams represent the average expression level of *OsPPPDE* genes in various tissues of rice, with red indicating a higher expression level and blue indicating a lower expression level. (B) Multiconditional expression level of *OsPPPDE* genes. Color annotations are the same as before. (C) Schematic representation of the *OsPPPDE* co-expression network in the normal treatment leaves (blue) and leaves after blast inoculation (red). Colored ovals indicate the number of co-expression pairs detected in normal leaves and leaves inoculated with blast.

### 3.5. Domestication of OsPPPDEs in Rice

To characterize the genetic variation of the *OsPPPDE* genes, 734 SNPs were identified in the 3000 Rice Genome Project [43]. As shown in Table S3, all members of the *OsPPPDE* genes had more than 50 SNPs in the gene and flanking regions except for *OsPPPDE2*, of which the SNP number was 13. The number of SNPs in the flanking region, both the 5'-flanking region and 3'-flanking region, was larger than that in the gene region. The promoter of *OsPPPDE1* had the largest number of SNPs (59) compared to others, while *OsPPPDE4* had the largest number of SNPs (49) in the 3'-flanking region. *OsPPPDE2* and *OsPPPDE7* had three and five SNPs in the 3'-flanking region, respectively, of which the number of SNPs was much smaller than the others. We also analyzed the LD pattern using 734 SNPs (Figure S5). This result showed that the *OsPPPDE4* locus was a strong LD region in all subgroups, especially in *japonica* cultivars.

Moreover, we examined the evolutionary relationship by estimating the level of fixation index ( $F_{ST}$ ), nucleotide diversity ( $\pi$ ), and Tajima's  $D$  (Figure 6). Pairwise measurements of  $F_{ST}$  showed that population differences were higher between *japonica* and *indica* in the *OsPPPDE7* gene and flanking regions, but strikingly weaker differences were found at the *OsPPPDE1* and *OsPPPDE4* loci. The *OsPPPDE3*, *OsPPPDE5*, and *OsPPPDE10* gene regions had low mean  $F_{ST}$ . The nucleotide diversity of five groups (*japonica*, *indica*, admix, ecotypes from Aus, Boro, and Rayada and aromatic varieties from Basmati and Sadri) was calculated (Figure 6B). All groups exhibited similar patterns of nucleotide diversity in all *OsPPPDE* genes. Nucleotide diversity of the *indica* cultivars was higher than that of *japonica* in most *OsPPPDE* genes except for *OsPPPDE8*, and a low  $\pi$  was detected of *japonica* and *indica* in *OsPPPDE7* and *OsPPPDE9*. The trend in all *OsPPPDE* genes of Tajima's  $D$  was similar to that of  $\pi$  (Figure 6C). Signals of artificial selection were found in all *OsPPPDE* genes. For instance, these had Tajima's  $D$  of *OsPPPDE9*, which were positive in aus, bas, and admix accessions and negative in the *japonica* and *indica* cultivars.

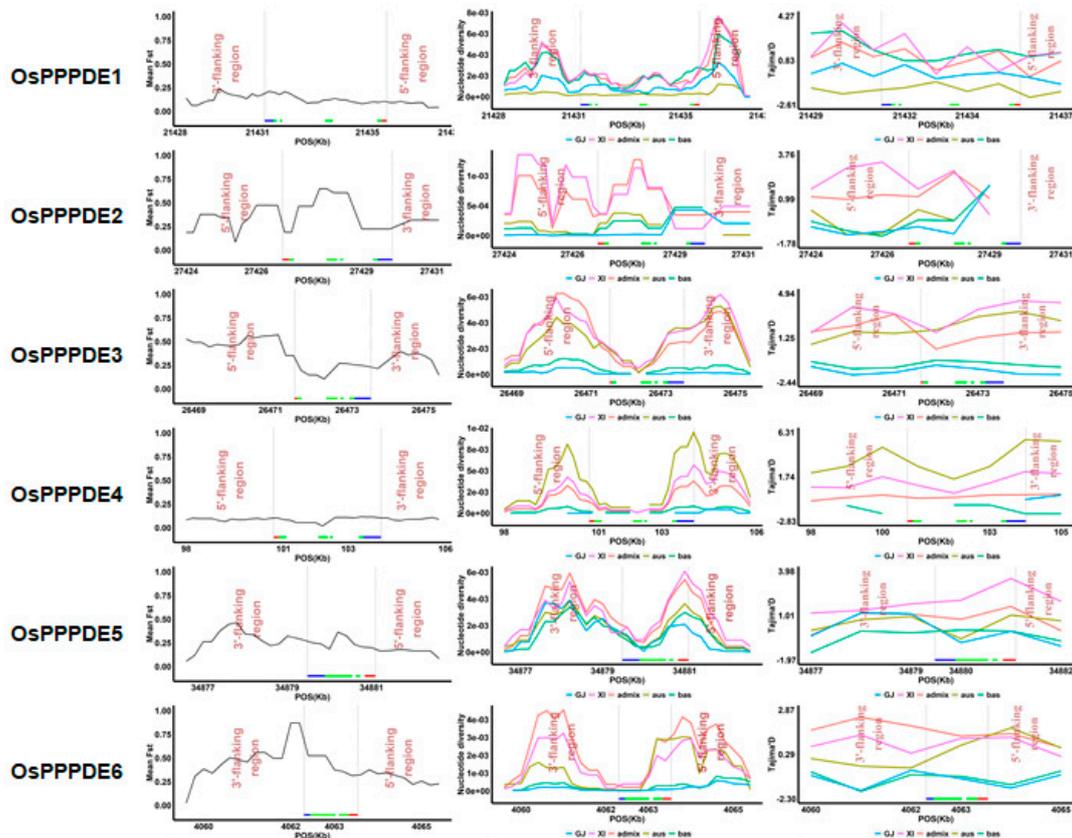
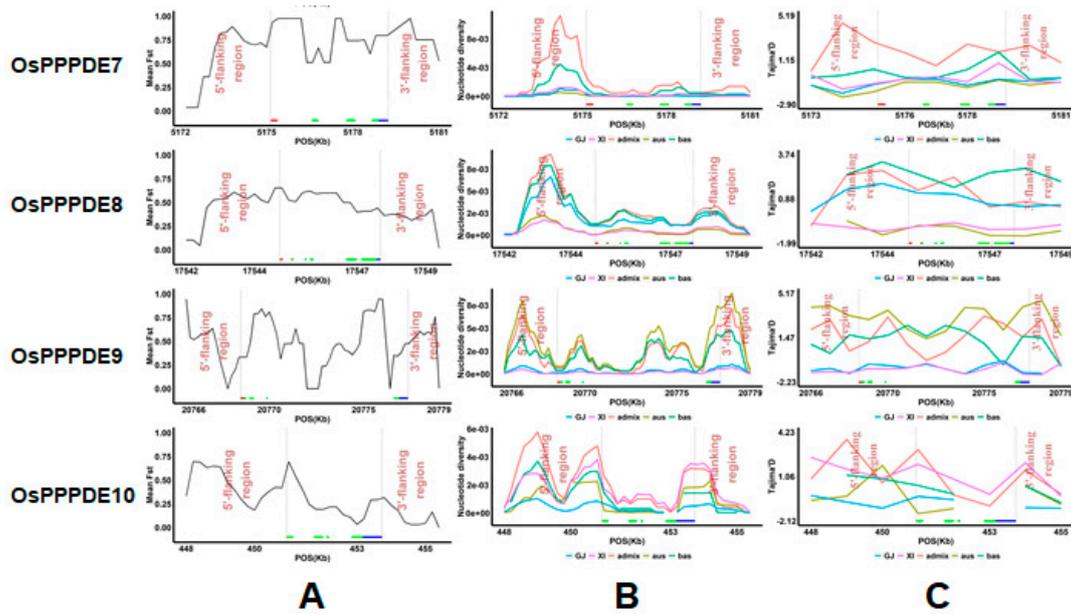


Figure 6. Cont.

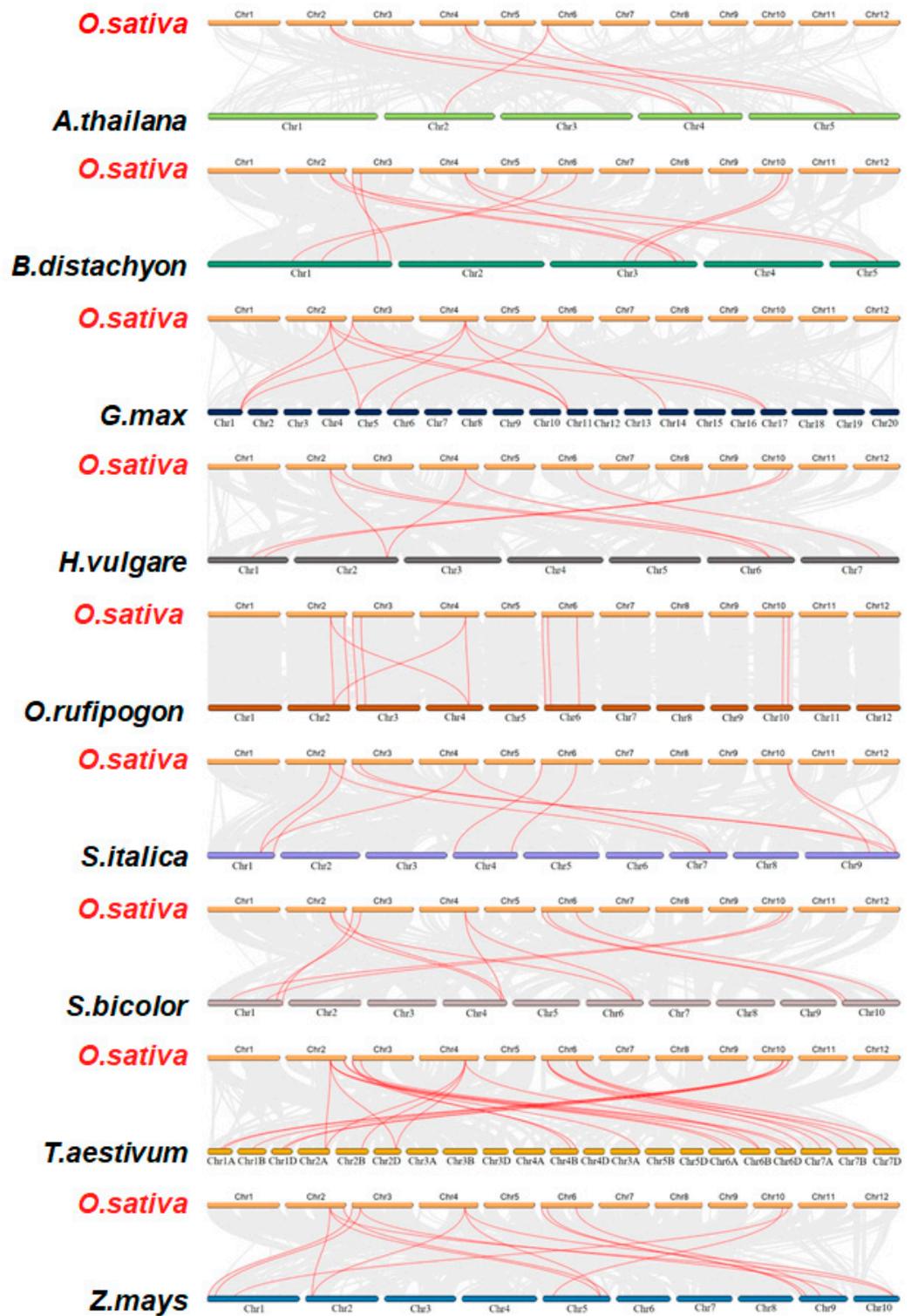


**Figure 6.** Natural variation in *OsPPPDE* genes contributes to subspecies divergence. (A)  $F_{ST}$  of *OsPPPDE* genes and flanking regions between *japonica* and *indica*. The line at the bottom of the figure shows the gene structure: red is 5'UTR; green is CDS; blue is 3'UTR. (B) Nucleotide diversity ( $\pi$ ) of *OsPPPDE* genes and flanking regions in *japonica* (GJ), *indica* (XI), and admix ecotypes from Aus, Boro, and Rayada (aus) and aromatic varieties from Basmati and Sadri (bas). (C) Tajima's  $D$  values of the *OsPPPDE* genes and flanking regions in 5 rice groups.

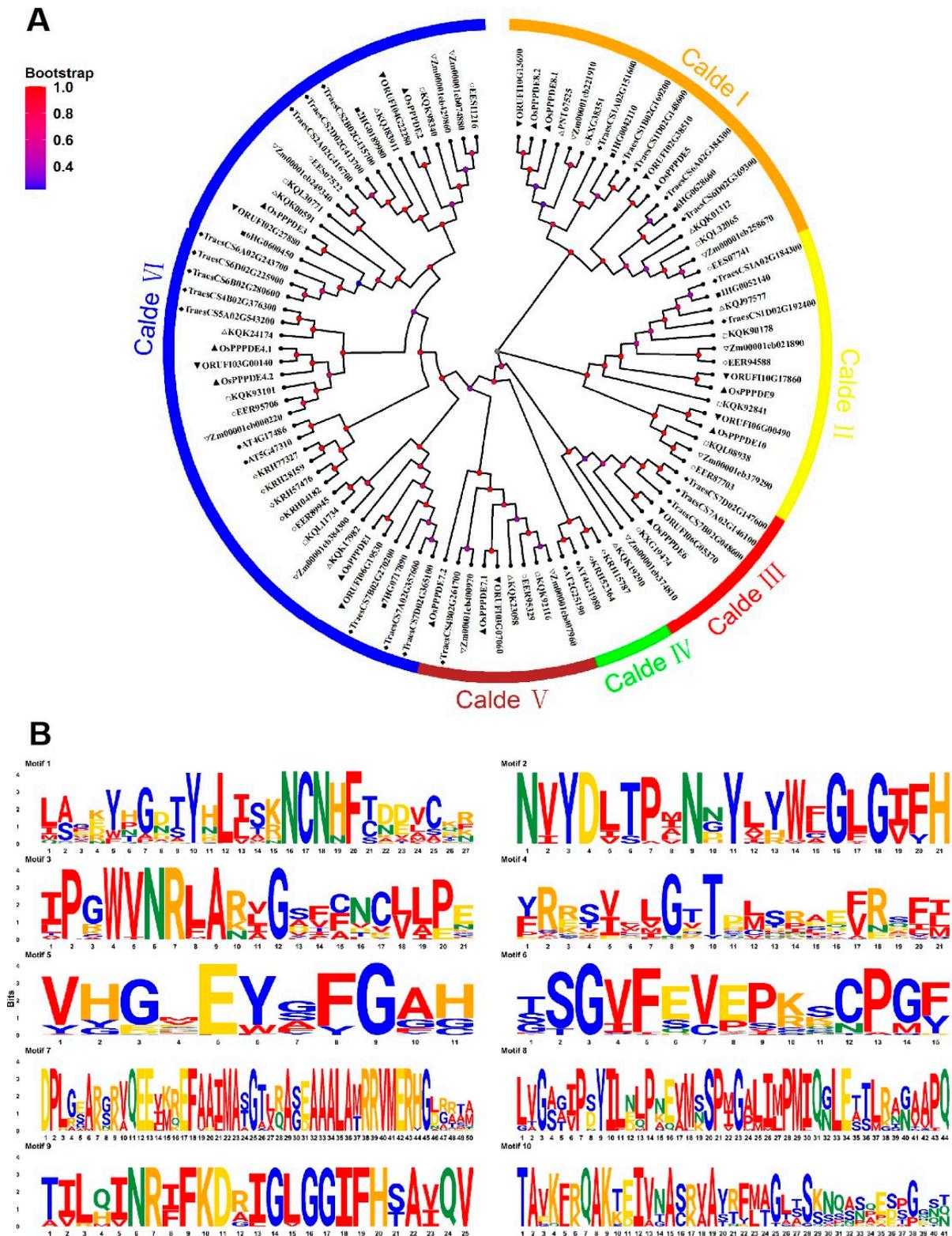
### 3.6. Synteny Analysis, Phylogenetic Analysis, and Motif Composition of the PPPDEs in Ten Plants

To clarify the evolutionary relationship of PPPDE proteins, we traced the syntenic relationship between *OsPPPDE* and homologs in seven monocotyledons (brachypodium, barley, wild rice, millet, sorghum, wheat, maize) and two dicotyledons (*Arabidopsis*, soybean). In the above species genomes, nine, six, ten, nine, ten, twenty-two, twelve, four, and six homologous genes were detected successively (Figure 7). The results indicated that the relationship between *OsPPPDEs* and PPPDEs in wheat, wild rice, maize, and sorghum was close. *OsPPPDE2*, *OsPPPDE3*, and *OsPPPDE6* showed the most homolog pairs with other plants. There were twelve, eight, five, and five homolog gene pairs identified between rice and wheat, maize, wild rice, and sorghum, respectively. These results imply that these genes may be conserved during evolution in the PPPDE gene family.

To further study the evolutionary diversification of PPPDE, we also constructed a neighbor-joining (NJ) tree of 101 PPPDE full length amino acid sequences from ten species using MEGA-X. All members of PPPDE could be divided into six groups labeled clade 1 to 6 (Figure 8A). In detail, *OsPPPDE8.1*, *OsPPPDE8.2*, and *OsPPPDE5* were in clade 1, *OsPPPDE9* and *OsPPPDE10* were in clade 2, *OsPPPDE6* was in clade 3, *OsPPPDE7.1* and *OsPPPDE7.2* were in clade 5, and *OsPPPDE1*, *OsPPPDE2*, *OsPPPDE4.1*, and *OsFTIP4.2* were in clade 6. We also observed that compared to other monocots, in all clades, wild rice PPPDEs were closer to rice PPPDE proteins. These results indicate that *OsPPPDEs* originated from wild rice, and there were no obvious genetic differentiations.



**Figure 7.** Syntenic analysis of PPPDE genes between rice and nine representative plant species. Gray line represents the collinear pair of genes of rice, the red line shows the collinear pairs of PPPDE genes. The gray lines at the bottom indicate the collinear blocks within the rice and other plant genomes. The red lines indicate the pairs of PPPDE genes. The results of the syntenic analysis between rice and the model plants including Arabidopsis, brachypodium, soybean, barley, wild rice, millet, sorghum, wheat, and maize.



**Figure 8.** Phylogenetic analysis of PPPDE proteins. (A) Evolution relationships among ten plant species. The tree was constructed with MEGA-X software using the neighbor-joining method (bootstrap = 1000 replicates). The tree divided the PPPDE proteins into 6 classes. (B) Conserved motif sequences in the PPPDE proteins of plants. MEME database was used to identify the motifs.

We used the MEME program to predict 10 conserved motifs in 101 PPPDE proteins, and these predicted motifs were annotated with the InterProScan program (Figures 8B and S2).

Based on the motif analysis, all of the PPPDE proteins contained motifs 1, 3, 4, and 5, except for clade 1. Moreover, the protein in clade 1 included motif 7 and motif 8, while the others did not. Clade 2 also detected subgroup-specific motifs: motif 9 and 10. Motif 1, motif 7, and motif 8 were related to the PPPDE domain, but the others had no functional annotation. In addition, the orthologous gene showed similar motif compositions, indicating that there were no functional differences among them.

#### 4. Discussion

The PPPDE superfamily, which is conserved in eukaryotes including humans, is a deubiquitinase (DUB). Members of this family have been identified as putative deubiquitinating isopeptidases and are involved in deubiquitination and/or deSUMOylation in mammals [21]. In previous research, PPPDE1 was overexpressed in hepatocellular carcinoma (HCC), which is a key modulator of the p53 protein and its downstream pathway [22]. However, the function of PPPDEs has been reported in Arabidopsis and rice, and the biological information in most crop species remains uncertain. *AtC3H59* regulates cell division by interacting with the PPPDE family protein Desi1 through its WD40 domain [23]. The deubiquitinase PIC1 contains a PPPDE domain as an immunity hub for PTI and ETI in rice [24]. The PPPDE structure domain, therefore, may be involved in plant cell division and the immune response. In this study, ten members in OsPPPDE family were characterized.

##### 4.1. The Gene Structure and Domains in OsPPPDEs Are Conserved

The gene structure analysis of the *OsPPPDEs* disclosed that these genes had multiple introns, with two to five (Figure 2B). However, another study asserted that plants retained intronless or less introns genes to manage stress conditions during evolution [48]. In our study, there were three genes that had intron retention due to alternative splicing. Intron retention also plays a significant part in the post-transcription of a variety of stressful conditions [49]. The PPPDE domains in PPPDE1 and OsPPPDE9 were mediated by the communication with other factors including p53 in humans and OsMETS [22,24]. PPPDE1 and OsPPPDE9 play an essential role in the ubiquitin signaling network. Only the amino acid sequences of the PPPDE domain were conserved among the rice PPPDE proteins (Figure 2B and Figure S3), and these results provide clues for analyzing the functions of other members of the OsPPPDE family. We found that there were collinearity relationships between *OsPPPDE2* and *OsPPPDE3*, which means that the OsPPPDEs family has experienced gene expansion events driven by segmental duplication. Furthermore, phylogenetic analysis categorized all OsPPPDE proteins into three distinct lineages differing from the exon–intron association and motif arrangement (Figure 2). Both group I and II members in rice were the ethylene-responsive element-binding protein.

##### 4.2. OsPPPDEs Are Universally Induced by Biotic and Abiotic Stress

Through the gene expression profiling of 352 rice samples from 30 projects, it was found that most *OsPPPDE* family members were differentially expressed in different treatments and regions (Figure 5). For example, *OsPPPDE5* is highly induced in dry leaves and cadmium-treated roots [50]. *OsPPPDE5* is a thioredoxin (TRX) family protein. The redox regulation of TRX is a specific control point for signal transduction pathways related to plant growth and stress response [51]. Therefore, *OsPPPDE5* may play an important role in response to drought and cadmium stress.

Gene expression is regulated by the complex interaction of many cis-acting elements and trans-acting factors involved in various pathways [52]. Notably, promoter analysis of the *OsPPPDE* gene revealed the presence of various stress response elements such as light response, phytohormone response, cold response, defense and stress responses, and anaerobic inducible elements (Figure 3). In a previous study, *OsPPPDE9* played a key role in broad-spectrum blast resistance mediated by the nucleotide binding domain and leucine-rich-repeat containing receptors (NLRs). The model of *OsPPPDE9* deubiquitinate

OsMETS mediates the methionine-ethylene cascade used in both PTI and ETI, which is ETI-PTI. Meanwhile, NLRs in the plant immune system such as PigmR protect *OsPPPDE9* from effector-mediated degradation [24]. *PPPDEs* are likely to respond to abiotic and biological stresses by deubiquitination. The seven *OsPPPDE* promoters also included multiple MYB binding sites (MBS), suggesting that some of these may be regulated by MYB transcription factors. In this study, multiple *OsPPPDEs* were found to be upregulated under drought stress conditions (Figure 5B). *OsPPPDE5*, a thioredoxin protein, was significantly upregulated under Ca treatment, while the others were not. Pleiotropic effects of *PPPDE* on seed germination, seedling development, and seed development were found in previous studies [23]. Furthermore, we found that some meristem expression and cell cycle regulation elements were also found in some *OsPPPDE* gene promoters (Figure 3). The expression level of rice *PPPDE* was significantly different in different tissues and developmental stages. This suggests that the *OsPPPDE* gene may be involved in plant growth and cell differentiation.

#### 4.3. Evolutionary Conservation of *PPPDEs* in Crops

Collinear analysis of *PPPDEs* and *PPPDE*-like proteins indicated that the number of *PPPDE* gene families was evolutionarily conserved, as many *PPPDE* orthologous pairs existed in the other nine crops and were similar in both monocotyledonous and dicotyledonous plants. The most orthologous pairs were found in wheat, with 22 pairs, suggesting that whole-genome duplication is extremely important for the expansion of *PPPDE* gene family members [53]. The researchers analyzed the *PPPDE* genes of Arabidopsis and found that *AtPPPDEs* can be divided into three groups, similar to our results [23]. In a multispecies phylogenetic tree, genes in a subgroup often have similar functions. In these studies, phylogenetic analyses of the *PPPDE* family in rice and other species were performed using their amino acid sequences. The results showed that the monocotyledonous and dicotyledonous orthologs clustered in the same clade (Figure 8A). MEME analysis revealed that 10 conserved motifs were identified in the *PPPDE* protein. The structures of *PPPDE* proteins in the phylogenetic group had similar motifs (Figure S2). Previous studies revealed that genes typically undergo tandem repeat events to amplify gene family members during the evolutionary process [54]. In this study, gene duplication was identified between *OsPPPDE2* and *OsPPPDE3* (Figure S1) as these two genes had collinear gene pairs with the other nine species (Figure 7). These results suggest that *PPPDEs* amplify gene family members by tandem repeat events in most crops, and that the *PPPDE* genes are more primitive than the divergence between monocots and dicots.

#### 4.4. *OsPPPDEs* Alleles Confer *Indica*–*Japonica* Divergence

Asian cultivated rice is thought to have been domesticated from wild rice species such as *Oryza rufipogon* and *Oryza nivara* [43,55]. The genome of wild rice is more diverse than cultivated rice, suggesting that various wild rice species are valuable resources for rice genetic improvement. *OsPPPDEs* showed significant collinearity with those in wild rice. The amount and chromosomal location of the *PPPDE* protein in wild rice were similar to those in cultivated rice. All *OsPPPDEs* paired with homologous chromosomes in wild rice (Figure 7). One study showed that *PIC11*, a member of the rice *PPPDE* gene family, plays a role in *indica*–*japonica* domestication, and further found that the promoter of *PIC11* in *japonica* had a higher transcriptional activity than *indica* after chitin treatment [24]. Our results suggest that selection signals are present in most members of the *OsPPPDE* family (Figure 6), indicating that *OsPPPDEs* play a role in *indica*–*japonica* differentiation.

From the 3K Rice Genome Project and Rice Expression Database and GEO database, 10 *OsPPPDE* genes were identified and characterized using a variety of biological information software. *OsPPPDE5* was related to drought tolerance, Cd and rice blast stress showed upregulated expression under comprehensive analysis (Figure 3A). Furthermore, three genes including *OsPPPDE3/6/9* were downregulated, indicating that these genes may negatively regulate abiotic and biotic stresses. Taken together, these results provide

valuable information in understanding the rice PPPDE gene family. It follows that we plan to combine a variety of rice SNPs datasets and use wet lab experiments to validate the PPPDE family functions to determine the deubiquitination function of various haplotypes under abiotic and biotic stresses.

## 5. Conclusions

This study investigated the genome-wide identification, structure, expression pattern, domestication, and evolution analysis of PPPDE genes in rice. The motifs of OsPPPDEs were highly conserved throughout the evolutionary history of rice. Most *OsPPPDE* genes were highly expressed in the panicle, roots, and leaves, and *OsPPPDE5* may be involved in the response to drought and cadmium stresses. Selection signals were present in most members of the *OsPPPDE* family, suggesting that *OsPPPDEs* may play a role in *indica-japonica* differentiation.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agronomy14051035/s1>, Figure S1: Syntenic analysis of *OsPPPDE* genes; Figure S2: MEME domain analysis of PPPDE genes in different plants; Figure S3: The 3D structure modeling of *OsPPPDE* proteins and PPPDE domain; Figure S4: The expression levels of the rice PPPDE genes; Figure S5: LD heat map for *OsPPPDE* gene and flanking regions in 5 rice groups; Table S1: The detailed experiment information from RED and GEO datasets; Table S2: The expression abundance data without the batch effect; Table S3: SNPs of *OsPPPDE* genes and flanking regions.

**Author Contributions:** Conceptualization, X.L. and Y.H.; methodology, W.L. and Y.H.; software, W.L.; writing—original draft preparation, W.L. and Y.H.; writing—review and editing, X.L. and Y.H.; supervision, D.C., W.W., Q.L. and Y.Z.; project administration, S.C. and L.C.; funding acquisition, X.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** The project was supported by the Nanfan special project, CAAS, (Grant Number: ZDXM2317; Grant Number YDLH2302), the China Agriculture Research System (Grant Number CARS-01-03), the Zhejiang Provincial Key Special Projects (Grant Number 2021C02063-1), and the Major Special Project of Guangxi Science and Technology (Grant Number AA23062015). We also acknowledge the Natural Science Foundation of China (31961143016, 31701338), the National Key Research and Development Program (2018YFD0100806), and the Strategic Science and Technology Innovation Cooperation Project of the Ministry of Agriculture and Rural Affairs (2020FYE0202300).

**Data Availability Statement:** Data is contained within the article.

**Conflicts of Interest:** The authors declare no conflicts of interests and that our research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

## References

- Schauer, N.J.; Magin, R.S.; Liu, X.; Doherty, L.M.; Buhrlage, S.J. Advances in Discovering Deubiquitinating Enzyme (DUB) Inhibitors. *J. Med. Chem.* **2020**, *63*, 2731–2750. [CrossRef]
- Luo, Y.; Takagi, J.; Claus, L.A.N.; Zhang, C.; Yasuda, S.; Hasegawa, Y.; Yamaguchi, J.; Shan, L.; Russinova, E.; Sato, T. Deubiquitinating enzymes UBP12 and UBP13 stabilize the brassinosteroid receptor BRI1. *EMBO Rep.* **2022**, *23*, e53354. [CrossRef]
- March, E.; Farrona, S. Plant Deubiquitinases and Their Role in the Control of Gene Expression Through Modification of Histones. *Front. Plant Sci.* **2017**, *8*, 2274. [CrossRef]
- Tang, X.; Ghimire, S.; Liu, W.; Fu, X.; Zhang, H.; Sun, F.; Zhang, N.; Si, H. Genome-wide identification of U-box genes and protein ubiquitination under PEG-induced drought stress in potato. *Physiol. Plant.* **2021**, *174*, e13475. [CrossRef]
- Nakagawa, T.; Nakayama, K. Protein monoubiquitylation: Targets and diverse functions. *Genes. Cells.* **2015**, *20*, 543–562. [CrossRef]
- Yang, J.; Chen, D.; Matar, K.A.O.; Zheng, T.; Zhao, Q.; Xie, Y.; Gao, X.; Li, M.; Wang, B.; Lu, G.D. The deubiquitinating enzyme MoUbp8 is required for infection-related development, pathogenicity, and carbon catabolite repression in *Magnaporthe oryzae*. *Appl. Microbiol. Biotechnol.* **2020**, *104*, 5081–5094. [CrossRef]
- Chung, C.H.; Baek, S.H. Deubiquitinating Enzymes: Their Diversity and Emerging Roles. *Biochem. Biophys. Res. Commun.* **2000**, *266*, 633–640. [CrossRef]
- Amerik, A.Y.; Hochstrasser, M. Mechanism and function of deubiquitinating enzymes. *Biochim. Biophys. Acta* **2004**, *1695*, 189–207. [CrossRef]

9. Taya, S.; Yamamoto, T.; Kanai-Azuma, M.; Wood, S.A.; Kaibuchi, K. The deubiquitinating enzyme Fam interacts with and stabilizes  $\beta$ -catenin. *Genes Cells* **1999**, *4*, 757–767. [[CrossRef](#)]
10. Luo, R.; Yang, K.; Xiao, W. Plant deubiquitinases: From structure and activity to biological functions. *Plant Cell Rep.* **2023**, *42*, 469–486. [[CrossRef](#)]
11. Makarova, K.S.; Aravind, L.; Koonin, E.V. A novel superfamily of predicted cysteine proteases from eukaryotes, viruses and *Chlamydia pneumoniae*. *Trends Biochem. Sci.* **2000**, *25*, 50–52. [[CrossRef](#)]
12. Makarova, K.S.; Aravind, L.; Koonin, E.V. A superfamily of archaeal, bacterial, and eukaryotic proteins homologous to animal transglutaminases. *Protein Sci.* **2010**, *8*, 1714–1719. [[CrossRef](#)]
13. Hartmut, S.; Stefan, T.; Kay, H. Elucidation of ataxin-3 and ataxin-7 function by integrative bioinformatics. *Hum. Mol. Genet.* **2003**, *12*, 2845–2852.
14. Vogel, K.; Isono, E. Deubiquitylating enzymes in *Arabidopsis thaliana* endocytic protein degradation. *Biochem. Soc. Trans.* **2024**, *52*, 291–299. [[CrossRef](#)]
15. Zhao, J.; Zhou, H.; Zhang, M.; Gao, Y.; Li, L.; Gao, Y.; Li, M.; Yang, Y.; Guo, Y.; Li, X. Ubiquitin-specific protease 24 negatively regulates abscisic acid signalling in *Arabidopsis thaliana*. *Plant Cell Environ.* **2016**, *39*, 427–440. [[CrossRef](#)]
16. Keren, I.; Citovsky, V. The histone deubiquitinase OTLD1 targets euchromatin to regulate plant growth. *Sci. Signal.* **2016**, *9*, ra125. [[CrossRef](#)]
17. Iyer, L.M.; Koonin, E.V.; Aravind, L. Novel predicted peptidases with a potential role in the ubiquitin signaling pathway. *Cell Cycle* **2004**, *3*, 1440–1450. [[CrossRef](#)]
18. Suh, H.Y.; Kim, J.H.; Woo, J.S.; Ku, B.; Shin, E.J.; Yun, Y.; Oh, B.H. Crystal structure of DeSI-1, a novel deSUMOylase belonging to a putative isopeptidase superfamily. *Proteins* **2012**, *80*, 2099–2104. [[CrossRef](#)]
19. Shin, E.J.; Shin, H.M.; Nam, E.; Kim, W.S.; Kim, J.H.; Oh, B.H.; Yun, Y. DeSUMOylating isopeptidase: A second class of SUMO protease. *EMBO Rep.* **2012**, *13*, 339–346. [[CrossRef](#)]
20. He, Y.; Wang, J.; Gou, L.; Shen, C.; Chen, L.; Yi, C.; Wei, X.; Yang, J. Comprehensive analysis of expression profile reveals the ubiquitous distribution of PPPDE peptidase domain 1, a Golgi apparatus component, and its implications in clinical cancer. *Biochimie* **2013**, *95*, 1466–1475. [[CrossRef](#)]
21. Xie, X.; Wang, X.; Jiang, D.; Wang, J.; Fei, R.; Cong, X.; Wei, L.; Wang, Y.; Chen, H. PPPDE1 is a novel deubiquitinase belonging to a cysteine isopeptidase family. *Biochem. Biophys. Res. Commun.* **2017**, *488*, 291–296. [[CrossRef](#)]
22. Xie, X.; Wang, X.; Liao, W.; Fei, R.; Wu, N.; Cong, X.; Chen, Q.; Wei, L.; Wang, Y.; Chen, H. PPPDE1 promotes hepatocellular carcinoma development by negatively regulate p53 and apoptosis. *Apoptosis* **2019**, *24*, 135–144. [[CrossRef](#)]
23. Seok, H.Y.; Bae, H.; Kim, T.; Mehdi, S.M.M.; Nguyen, L.V.; Lee, S.Y.; Moon, Y.H. Non-TZF Protein AtC3H59/ZFWD3 Is Involved in Seed Germination, Seedling Development, and Seed Development, Interacting with PPPDE Family Protein Desi1 in *Arabidopsis*. *Int. J. Mol. Sci.* **2021**, *22*, 4738. [[CrossRef](#)]
24. Zhai, K.; Liang, D.; Li, H.; Jiao, F.; Yan, B.; Liu, J.; Lei, Z.; Huang, L.; Gong, X.; Wang, X.; et al. NLRs guard metabolism to coordinate pattern- and effector-triggered immunity. *Nature* **2021**, *601*, 245–251. [[CrossRef](#)]
25. Kawahara, Y.; Bastide, M.; Hamilton, J.P.; Kanamori, Z.; Huang, L.; Gong, X.; Wang, X. NLRs genome using next generation sequence and optical map data. *Rice* **2013**, *6*, 4. [[CrossRef](#)]
26. Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G.A.; Sonnhammer, E.L.L.; Tosatto, S.C.E.; Paladin, L.; Raj, S.; Richardson, L.J.; et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* **2021**, *49*, D412–D419. [[CrossRef](#)]
27. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [[CrossRef](#)]
28. Letunic, I.; Khedkar, S.; Bork, P. SMART: Recent updates, new developments and status in 2020. *Nucleic Acids Res.* **2021**, *49*, D458–D460. [[CrossRef](#)]
29. Wilkins, M.R.; Gasteiger, E.; Bairoch, A.; Sanchez, J.C.; Hochstrasser, D.F. Protein Identification and Analysis Tools in the ExPASy Server. *Methods Mol. Biol.* **1999**, *112*, 531–552.
30. Horton, P.; Park, K.J.; Obayashi, T.; Fujita, N.; Harada, H.; Adams-Collier, C.J.; Nakai, K. WoLF PSORT: Protein localization predictor. *Nucleic Acids Res.* **2007**, *35*, W585–W587. [[CrossRef](#)]
31. Wang, Y.; Tang, H.; Debarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [[CrossRef](#)]
32. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant.* **2020**, *13*, 1194–1202. [[CrossRef](#)]
33. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [[CrossRef](#)]
34. Yu, G. Using ggtree to Visualize Data on Tree-Like Structures. *Curr. Protoc. Bioinform.* **2020**, *69*, e96. [[CrossRef](#)]
35. Bailey, T.L.; Johnson, J.; Grant, C.E.; Noble, W.S. The MEME Suite. *Nucleic Acids Res.* **2015**, *43*, W39–W49. [[CrossRef](#)]
36. Lescot, M. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **2002**, *30*, 325–327. [[CrossRef](#)]
37. Ginestet, C. ggplot2: Elegant Graphics for Data Analysis. *J. R. Stat. Soc.* **2011**, *174*, 245–246. [[CrossRef](#)]
38. Geourjon, C.; Deleage, G. SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.* **1995**, *11*, 681–684.

39. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **2022**, *50*, D439–D444. [[CrossRef](#)]
40. Delano, W.L. The PyMol Molecular Graphics System. *Proteins Struct. Funct. Bioinform.* **2002**, *30*, 442–454.
41. Xia, L.; Zou, D.; Sang, J.; Xu, X.; Yin, H.; Li, M.; Wu, S.; Hu, S.; Hao, L.; Zhang, Z. Rice Expression Database (RED): An integrated RNA-Seq-derived gene expression database for rice. *J. Genet. Genomics.* **2017**, *44*, 235–241. [[CrossRef](#)]
42. Maag, J.L.V. gganatogram: An R package for modular visualisation of anatograms and tissues based on ggplot2. *F1000Research* **2018**, *7*, 1576. [[CrossRef](#)]
43. Wang, W.; Mauleon, R.; Hu, Z.; Chebotarov, D.; Tai, S.; Wu, Z.; Li, M.; Zheng, T.; Fuentes, R.R.; Zhang, F.; et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **2018**, *557*, 43–49. [[CrossRef](#)]
44. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [[CrossRef](#)]
45. Shin, J.H.; Blay, S.; Mcnoney, B.; Graham, J. LDheatmap: An R Function for Graphical Display of Pairwise Linkage Disequilibrium between Single Nucleotide Polymorphisms. *J. Stat. Softw.* **2006**, *16*, 3. [[CrossRef](#)]
46. Soltani, B.M.; Ehlting, J.; Hamberger, B.; Douglas, C.J. Multiple cis-regulatory elements regulate distinct and complex patterns of developmental and wound-induced expression of Arabidopsis thaliana 4CL gene family members. *Planta* **2006**, *224*, 1226–1238. [[CrossRef](#)]
47. Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Zidek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; et al. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*, 590–596. [[CrossRef](#)]
48. Mattick, J.S.; Gagen, M.J. The Evolution of Controlled Multitasked Gene Networks: The Role of Introns and Other Noncoding RNAs in the Development of Complex Organisms. *Mol. Biol. Evol.* **2001**, *18*, 1611–1630. [[CrossRef](#)]
49. Kannan, S.; Halter, G.; Renner, T.; Waters, E.R. Patterns of alternative splicing vary between species during heat stress. *AoB Plants* **2018**, *10*, ply013. [[CrossRef](#)]
50. Oono, Y.; Yazawa, T.; Kawahara, Y.; Kanamori, H.; Kobayashi, F.; Sasaki, H.; Mori, S.; Wu, J.; Handa, H.; Itoh, T.; et al. Genome-wide transcriptome analysis reveals that cadmium stress signaling controls the expression of genes in drought stress signal pathways in rice. *PLoS ONE* **2014**, *9*, e96946. [[CrossRef](#)]
51. Jiménez, A.; López-Martínez, R.; Martí, M.C.; Cano-Yelo, D.; Sevilla, F. The integration of TRX/GRX systems and phytohormonal signalling pathways in plant stress and development. *Plant Physiol. Biochem.* **2024**, *207*, 108298.
52. Zhang, Y.; Zheng, L.; Yun, L.; Ji, L.; Li, G.; Ji, M.; Shi, Y.; Zheng, X. Catalase (CAT) Gene Family in Wheat (*Triticum aestivum* L.): Evolution, Expression Pattern and Function Analysis. *Int. J. Mol. Sci.* **2022**, *23*, 542. [[CrossRef](#)]
53. Freeling, M. Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* **2009**, *60*, 433–453. [[CrossRef](#)]
54. Cannon, S.B.; Mitra, A.; Baumgarten, A.; Young, N.D.; May, G. The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. *BMC Plant Biol.* **2004**, *4*, 10. [[CrossRef](#)]
55. Kovach, M.J.; Sweeney, M.T.; McCouch, S.R. New insights into the history of rice domestication. *Trends Genet.* **2007**, *23*, 578–587. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.